

Home

Insert

Draw

View

Help



100 Days of ML



Day 14 - Framing L...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discretiza...

+ Add section

+ Add page

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

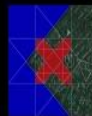
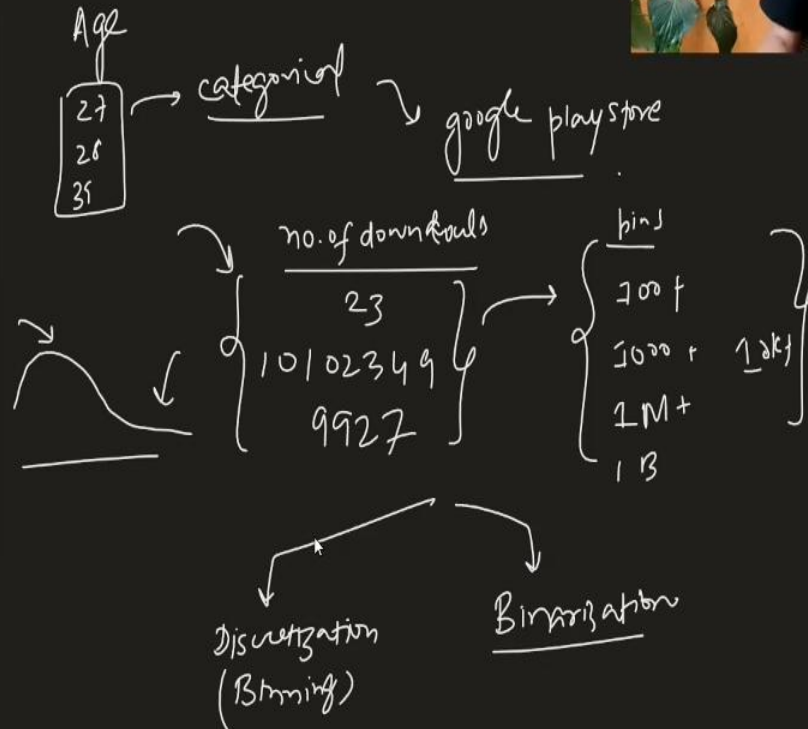
10. Binarization

11. Example

1. Encoding Numerical Features

Monday, April 19, 2021

4:18 PM





100 Days of ML



Day 14 - Framing L...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discretiza...

+ Add section

+ Add page

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

10. Binarization

11. Example

2. Discretization

Discretization is the process of transforming continuous variables into discrete variables by creating a set of contiguous intervals that span the range of the variable's values. Discretization is also called binning, where bin is an alternative name for interval.

Why use Discretization:

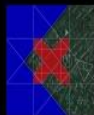
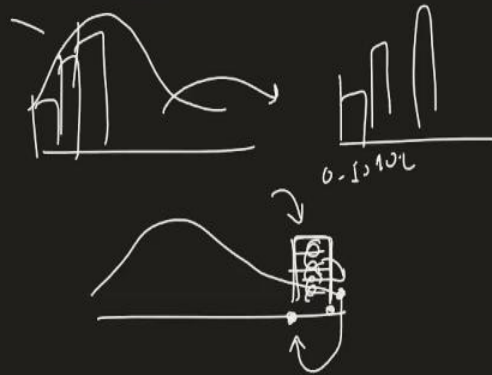
1. To handle Outliers
2. To improve the value spread

Age

23 42, 57 81, . . . 110

0-10, 10-20, 20-30, . . .

5 6 10



OneNote for Windows 10

Home Insert Draw View Help

100 Days of ML

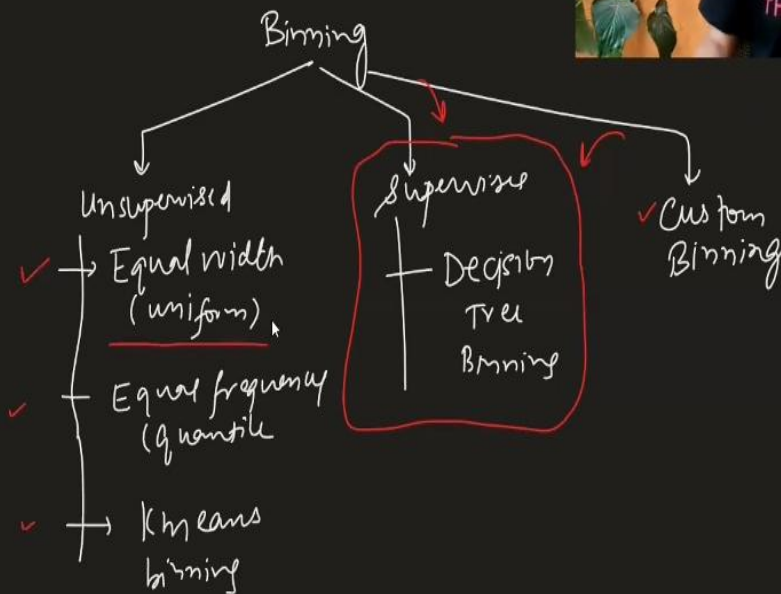
Day 14 - Framing L...
Day 15 - Working...
Day 16 - Working...
Day 17 - API to Pa...
Day 18 - Web Scra...
Day 19 - Understa...
Day 20 - Univariate...
Day 21 - Bivariate...
Day 22 - Pandas Pr...
Day 23 - Feature E...
Day 24 - Standardi...
Day 25 - Normaliza...
Day 26 - Ordinal E...
Day 27 - One Hot E...
Day 28 - Column Tr...
Day 29 - Pipelines
Day 30 - Function T...
Day 31 - Power Tra...
Day 32 - Discretiza...

1. Encoding Numerical Features
2. Discretization
3. Types of Discretization
4. Equal Width/Uniform Binning
5. Equal Frequency/Quantile Binning
6. KMeans Binning
7. Encoding the discretized variable
8. Example
9. Custom/Domain Based Binning
10. Binarization
11. Example

+ Add section + Add page

3. Types of Discretization

Monday, April 19, 2021 3:56 PM



Home

Insert

Draw

View

Help



Shapes

Link to Shape

Link to Text



100 Days of ML



Day 14 - Framing L...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discrtizati...

Day 33 - Discrtizati...

+ Add section

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

10. Binarization

11. Example

12. Example

13. Example

14. Example

15. Example

16. Example

17. Example

18. Example

19. Example

20. Example

21. Example

4. Equal Width/Uniform Binning

Monday, April 19, 2021 3:56 PM

Age

(27), 32, 84, 56, ... max 100

Bins = 10 min 0

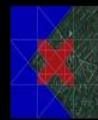
$$\frac{\text{max} - \text{min}}{\text{bins}} = \frac{100 - 0}{10} = 10$$

→ 10 bins

(0-10), (10-20), (20-30), ... (90-100)



	age	age_trf	age_labels
314	43.0	5.0	(40.21, 48.168]
523	44.0	5.0	(40.21, 48.168]
352	15.0	1.0	(8.378, 16.336]
534	30.0	3.0	(24.294, 32.252]
211	35.0	4.0	(32.252, 40.21]
530	2.0	0.0	(0.42, 8.378]
786	18.0	2.0	(16.336, 24.294]
827	1.0	0.0	(0.42, 8.378]
372	19.0	2.0	(16.336, 24.294]
518	36.0	4.0	(32.252, 40.21]



Home

Insert

Draw

View

Help





























































































































































































































































































































































Home

Insert

Draw

View

Help



100 Days of ML



Day 14 - Framing L...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discretiza...

Day 33 - Discretiza...

Day 34 - Discretiza...

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

5. Equal Frequency/Quantile Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

10. Binarization

11. Example

7. Encoding the discretized variable

Monday, April 19, 2021 3:58 PM

SKlearn

KMeansDiscretizer()

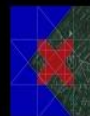
bins=?

strategy

+ uniform
+ quantile
+ kmeans

encoding

Calls and notifications will vibrate



Home

Insert

Draw

View

Help



100 Days of ML

Day 14 - Framing L...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discretiza...

+ Add section

+ Add page

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

10. Binarization

11. Example

10. Binarization

Monday, April 19, 2021 4:17 PM

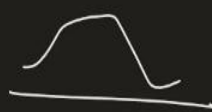


image 0,1

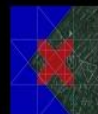
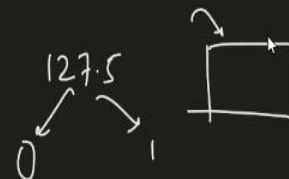


0-255
color

→ 0

Annual income

→ 62 <



Home

Insert

Draw

View

Help



100 Days of ML



Day 14 - Framing L...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discretiza...

+ Add section

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

10. Binarization

11. Example

+ Add page

7. Encoding the discretized variable

Monday, April 19, 2021 3:58 PM

SKlearn

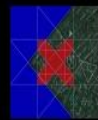
KMeansDiscretizer()

bins=?

strategy

+ uniform
+ quantile
+ kmeans

encoding

+ ordinal
+ onehotencoding

Home

Insert

Draw

View

Help



Shapes

Link to Shape

Link to Text



100 Days of ML



Day 14 - Framing L...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discretiza...

+ Add section

+ Add page

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

10. Binarization

11. Example

9. Custom/Domain Based Binning

Monday, April 19, 2021

3:58 PM

$$\left. \begin{array}{l} [0-18] \rightarrow \text{Kids} \\ [18-60] \rightarrow \\ [60-80] \rightarrow \end{array} \right\} \text{Stream} \rightarrow \text{Pandas}$$



Home

Insert

Draw

View

Help



Shapes

Link to Shape

Link to Text



100 Days of ML



Day 14 - Framing L...

Day 15 - Working...

Day 16 - Working...

Day 17 - API to Pa...

Day 18 - Web Scra...

Day 19 - Understa...

Day 20 - Univariate...

Day 21 - Bivariate...

Day 22 - Pandas Pr...

Day 23 - Feature E...

Day 24 - Standardi...

Day 25 - Normaliza...

Day 26 - Ordinal E...

Day 27 - One Hot E...

Day 28 - ColumnTr...

Day 29 - Pipelines

Day 30 - Function T...

Day 31 - Power Tra...

Day 32 - Discretiza...

+ Add section

1. Encoding Numerical Features

2. Discretization

3. Types of Discretization

4. Equal Width/Uniform Binning

5. Equal Frequency/Quantile Binni...

6. KMeans Binning

7. Encoding the discretized variable

8. Example

9. Custom/Domain Based Binning

10. Binarization

11. Example

10. Binarization

Monday, April 19, 2021 4:17 PM

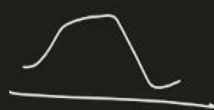


image 0,1

0-255
color

0

127.5
0

Annual income



62

